Chapter 5

Dynamic Panel Model

Objectives

(1) Introduce about Dynamic Panel Model

(2) Fixed and Random Effects Estimation

(3) Instrumental Variable Estimation (IV approach) (Anderson and Hsiao, 1982)

(4) 2SLS, Generalized Method of Moment (GMM) approach (Arenallo and Bond, 1985)

5.1 Introduction

Linear dynamic panel data models include lag dependent variables as covariates along with the unobserved effects, fixed or random, and exogenous regressor

$$y_{it} = \gamma_0 + \sum_{j=1}^{p} \gamma_j y_{t-j} + x_{it}\beta + \alpha_i + u_{it} = \sum_{j=1}^{p} \gamma_j y_{t-j} + x_{it}\beta + \alpha_i^* + u_{it} \qquad (5.1)$$

Notes: *The presence of lagged dependent variable as a regressor incorporates the entire history of it, and any impact of x_{it} on y_{it} is conditioned on this history.*

We consider a dynamic panel model, in the sense that it contains (at least) one lagged variables. For simplicity, let us consider

$$y_{it} = \gamma_1 y_{it-1} + \beta'_{it} x_{it} + \alpha_i^* + u_{it} \qquad (5.2)$$

$$y_{it} = \gamma_1 y_{it-1} + \beta'_{it} x_{it} + \alpha_i^* + u_{it} \qquad (5.2)$$

Eq. (5.2) requires that $|\gamma| < 1$

$$y_{it} = \gamma_1 y_{it-1} + \alpha_i^* + u_{it} = \gamma_0 + \gamma_1 y_{it-1} + \alpha_i + u_{it} \qquad (5.3)$$

Assumptions on random disturbance are the following:

About α_i,

$$E(\alpha_i) = 0, \quad ,V(\alpha_i) = E(\alpha_i^2) = \sigma_\mu^2, \quad ,E(\alpha_i x_{it}) = 0, \quad ,E(\alpha_i \alpha_j) = 0$$

About u_{it},

$$E(u_{it}) = 0, \quad ,V(u_{it}) = E(u_{it}^2) = \sigma_u^2, \quad ,E(u_{it} u_{js}) = 0 \text{ for } i \neq j \text{ and } t \neq s$$

$$E(u_{it} / y_{it-1}) = 0$$

$$E(\alpha_i / y_{it-1}) \neq 0$$

By setting t = 1, 2,… and so on, the autoregressive process can be expressed in the following way:

$$y_{i(t=1)} = \gamma_0 + \gamma y_{i0} + \alpha_i + u_{i0}$$

$$y_{i2} = \gamma_0 + \gamma y_{i1} + \alpha_i + u_{i2} = \gamma_0 + \alpha_i + \gamma_1\left(\gamma_0 + \gamma_1 y_{i0} + \alpha_i + u_{i0}\right) + u_{i2}$$

$$= \gamma_0 + \gamma_0\gamma_1 + \alpha_i + \alpha_i\gamma_1 + \gamma_1^2 y_{i0} + \gamma_1 u_{i1} + u_{i2}$$

....

$$y_{it} = \gamma_0\left(1 + \gamma_1 + \ldots + \gamma_1^{t-1}\right) + \alpha_i\left(1 + \gamma_1 + \ldots + \gamma_1^{t-1}\right) + \gamma_1^t y_{i0} + \sum_{j=0}^{t-1}\gamma_1^j u_{i,t-j}$$

Or

$$y_{it} = \gamma_0\sum_{j=0}^{t-1}\gamma_1^j + \alpha_i\sum_{j=0}^{t-1}\gamma_1^j + \gamma_1^t y_{i0} + \sum_{j=0}^{t-1}\gamma_1^j u_{i,t-j}$$

Therefore

$$y_{it-1} = \gamma_0\sum_{j=0}^{t-2}\gamma_1^j + \alpha_i\sum_{j=0}^{t-2}\gamma_1^j + \gamma_1^{t-1} y_{i0} + \sum_{j=0}^{t-2}\gamma_1^j u_{i,t-1-j}$$

For large t,

$$E\left(y_{it}/\alpha_i\right)=\gamma_0\frac{1}{1-\gamma_1}+\alpha_i\frac{1}{1-\gamma_1}$$

$$V\left(y_{it}/\alpha_i\right)=\frac{\sigma_\varepsilon^2}{1-\gamma_1^2}$$

5.2 Fixed and Random Effects Estimation

$$y_{it}=\gamma_0+\gamma_1 y_{it-1}+\alpha_i+u_{it} \quad (5.3)$$

Remark: *One possible cause for biasedness is the presence of the unknown individual effects* α_i*, which creates a correlation between the explanatory variables and the residuals*

$$\left(y_{it}-\overline{y}_i\right)=\gamma_1\left(y_{it-1}-\overline{y}_{i,-1}\right)+u_{it}-\overline{u}_i$$

Notes: $\left(y_{it-1}-\overline{y}_{i,-1}\right)$ *will be correlated* $\left(u_{it}-\overline{u}_i\right)$

$$\left(y_{it}-\bar{y}_i\right)=\gamma_1\left(y_{it-1}-\underbrace{\bar{y}_{i,-1}}_{\textit{depen on past value of } u_{it}}\right)+u_{it}-\underbrace{\bar{u}_i}_{\textit{depen on past value of } u_{it}}$$

The within estimator or fix effects estimator is

$$\hat{\gamma}_{1FE}=\frac{\sum_{i=1}^{N}\sum_{t=1}^{T}\left(y_{it}-\bar{y}_i\right)\left(y_{it-1}-\bar{y}_{i,-1}\right)}{\sum_{i=1}^{N}\sum_{t=1}^{T}\left(y_{it-1}-\bar{y}_{i,-1}\right)^2}$$

$$=\gamma_1+\frac{\sum_{i=1}^{N}\sum_{t=1}^{T}\left(y_{it-1}-\bar{y}_{i,-1}\right)\left(u_{it}-\bar{u}_i\right)}{\sum_{i=1}^{N}\sum_{t=1}^{T}\left(y_{it-1}-\bar{y}_{i,-1}\right)^2}$$

$$\hat{\alpha}_i=y_i-\hat{\gamma}_{1FE}\,\bar{y}_{i,-1}$$

Problem: Fixed effects the within transformation and LSDV produce biased estimates

$$\hat{\gamma}_{1FE} = \gamma_1 + \frac{\sum_{i=1}^{N}\sum_{t=1}^{T}\left(y_{it-1} - \bar{y}_{i,-1}\right)\left(u_{it} - \bar{u}_i\right) / NT}{\sum_{i=1}^{N}\sum_{t=1}^{T}\left(y_{it-1} - \bar{y}_{i,-1}\right)^2 / NT}$$

Theorem. (Weak law of large numbers, Khinchine)

If $\{X_i\}$ for i=1,…, m is a sequence of i.i.d random variables with $E(X_i)$ $= \mu < \infty$, then the sample mean coverges in probability to μ:

$$\frac{1}{m}\sum_{i=1}^{m} X_i \xrightarrow{p} E\left(X_i\right) = \mu \Leftrightarrow p\lim_{m\to+\infty}\frac{1}{m}\sum_{i=1}^{m} X_i = E\left(X_i\right) = \mu$$

We have

$$\underset{N\to+\infty}{p\lim}\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(y_{it-1}-\bar{y}_{i,-1}\right)\left(u_{it}-\bar{u}_{i}\right)$$

$$=\underbrace{\underset{N\to+\infty}{p\lim}\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}y_{it-1}u_{it}}_{N_1}-\underbrace{\underset{N\to+\infty}{p\lim}\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}y_{it-1}\bar{u}_{i}}_{N_2}$$

$$-\underbrace{\underset{N\to+\infty}{p\lim}\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\bar{y}_{i,-1}u_{it}}_{N_3}+\underbrace{\underset{N\to+\infty}{p\lim}\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\bar{y}_{i,-1}\bar{u}_{i}}_{N_4}$$

$$N_1=\underset{N\to+\infty}{p\lim}\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}y_{it-1}u_{it}=E\left(y_{it-1}u_{it}\right)=0$$

$$N_2 = \underset{N\to+\infty}{p\lim}\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} y_{it-1}\bar{u}_i = \underset{N\to+\infty}{p\lim}\frac{1}{NT}\sum_{i=1}^{N}\bar{u}_i\sum_{t=1}^{T} y_{it-1}$$

$$= \underset{N\to+\infty}{p\lim}\frac{1}{NT}\sum_{i=1}^{N}\bar{u}_i T\bar{y}_{i,-1} = \underset{N\to+\infty}{p\lim}\frac{1}{N}\sum_{i=1}^{N}\bar{u}_i\bar{y}_{i,-1}$$

$$N_3 = \underset{N\to+\infty}{p\lim}\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\bar{y}_{i,-1}u_{it} = \underset{N\to+\infty}{p\lim}\frac{1}{NT}\sum_{i=1}^{N}\bar{y}_{i,-1}\sum_{t=1}^{T}u_{it} = \underset{N\to+\infty}{p\lim}\frac{1}{N}\sum_{i=1}^{N}\bar{y}_{i,-1}\bar{u}_i$$

$$N_4 = \underset{N\to+\infty}{p\lim}\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\bar{y}_{i,-1}\bar{u}_i = \underset{N\to+\infty}{p\lim}\frac{1}{NT}T\sum_{i=1}^{N}\bar{y}_{i,-1}\bar{u}_i = \underset{N\to+\infty}{p\lim}\frac{1}{N}\sum_{i=1}^{N}\bar{y}_{i,-1}\bar{u}_i$$

$$\underset{N\to+\infty}{p\lim}\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(y_{it-1}-\bar{y}_{i,-1}\right)\left(u_{it}-\bar{u}_i\right) = -\underset{N\to+\infty}{p\lim}\frac{1}{N}\sum_{i=1}^{N}\bar{u}_i\bar{y}_{i,-1}$$

$$0 - \underset{N\to+\infty}{p\lim}\frac{1}{N}\sum_{i=1}^{N}\bar{u}_i\bar{y}_{i,-1} - \underset{N\to+\infty}{p\lim}\frac{1}{N}\sum_{i=1}^{N}\bar{u}_i\bar{y}_{i,-1} + \underset{N\to+\infty}{p\lim}\frac{1}{N}\sum_{i=1}^{N}\bar{u}_i\bar{y}_{i,-1}$$

$$\hat{\gamma}_{1FE} = \gamma_1 + \frac{\sum_{i=1}^{N}\sum_{t=1}^{T}\left(y_{it-1} - \bar{y}_{i,-1}\right)\left(u_{it} - \bar{u}_i\right)/NT}{\sum_{i=1}^{N}\sum_{t=1}^{T}\left(y_{it-1} - \bar{y}_{i,-1}\right)^2/NT} = \gamma_1 - \underset{N\to+\infty}{p\lim}\frac{1}{N}\sum_{i=1}^{N}\bar{u}_i\bar{y}_{i,-1}$$

If this plim is not null, then the $\hat{\gamma}_{1,FE}$ estimator is biased when N tends to infinity and T is fixed

Fact. If T also tends to infinity, then the numerator converges to zero

Fact. The problem is more prominent in the random effects model. The lagged dependent variable is correlated with the compound disturbance in the model.

$$y_{it} = \gamma_1 y_{it-1} + \alpha_i^* + u_{it} = \gamma_0 + \gamma_1 y_{it-1} + \alpha_i + u_{it} \quad (5.3)$$

$$E\left(y_{it-1}\alpha_i^*\right) = E\left(\gamma_0\sum_{j=0}^{t-2}\gamma_1^j + \alpha_i\sum_{j=0}^{t-2}\gamma_1^j + \gamma_1^{t-1}y_{i0} + \sum_{j=0}^{t-2}\gamma_1^j u_{i,t-1-j}\right)\alpha_i^* \neq 0$$

Pre Example (3.1) With model

ROAA = f(L.ROAA, HHI, L_A, SIZE, ASSET_GRO, GDP, INF) + ε

5.3 Instrumental Variable Estimation

5.3.1 Define the endogeneity bias and the smearing effect.

Consider the (population) multiple linear regression model

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

- y is a Nx1 vector of observation for y_j , j= 1,…,N
- X is a NxK matrix of K explicative variables x_{jk} for k=1,…,K and j=1,…,N
- $\beta = (\beta_1\ \beta_2\ \ldots\ \beta_K)'$ is a Kx1 vector of parameters
- ε is a Nx1 vector of error terms ε_i with $V(\varepsilon/X) = \sigma^2 I_N$

Endogeneity we assume that the assumption A1 (exogeneity) is violated

$$E(\varepsilon/X) \neq 0$$

With

$$p\lim \frac{1}{N} X'\varepsilon = E\left(x_j \varepsilon_j\right) = \gamma \neq 0_{K\times 1}$$

Theorem (Bias of the OLS estimator) If the regressors are endogenous, i.e. E (ε/X) 6= 0, the OLS estimator of β is biased

$$E\left(\widehat{\beta}_{OLS} / X\right) \neq \beta$$

where β denotes the true value of the parameters. This bias is called the **endogeneity bias.**

Theorem (Inconsistency of the OLS estimator) If the regressors are endogenous with **plim** $\mathbf{N^{-1}X'\varepsilon = \gamma}$ the OLS estimator of β is inconsistent

$$p\lim \widehat{\beta}_{OLS} = \beta + Q^{-1}\gamma$$

$$where \;\; Q = p\lim N^{-1} X' X$$

Proof: Given the definition of the OLS estimator

$$\widehat{\beta}_{OLS} = (X'X)^{-1} X'y = (X'X)^{-1} X'(X\beta + \varepsilon)$$

$$= \beta + (X'X)^{-1} (X'\varepsilon)$$

$$We \;\; have$$

$$p\lim \widehat{\beta}_{OLS} = \beta + p\lim\left(\frac{1}{N} X'X\right)^{-1} \times p\lim\left(\frac{1}{N} X'\varepsilon\right)$$

$$= \beta + Q^{-1}\gamma \neq \beta$$

Notes.

- The implication is that even though only one of the variables in X is correlated with ε, **all of the elements** $\widehat{\beta}_{OLS}$ **of are inconsistent**, not just the estimator of the coefficient on the endogenous variable

Notes (cont.).

- This effects is called **smearing effect**: the inconsistency due to the endogeneity of the one variable is smeared across all of the least squares estimators

5.3.2 Instrustment variable

Definition. Consider a set of H variables $z_h \in R^N$ for h = 1, ..N. Denote Z the NxH matrix (z_1 ... z_H). These variables are called instruments or instrumental variables if they satisfy two properties:

(1) Exogeneity: They are uncorrelated with the disturbance.

$$E(\varepsilon/Z)= 0_{Nx1}$$

(2) Relevance: They are correlated with the independent variables, X

$$E(x_{jk}z_{jh}) \neq 0 \text{ for } h \in \{1, .., H\} \text{ and } k \in \{1, ..,K\}.$$

Assumptions: The instrumental variables satisfy the following properties.

Well behaved data:

$$\text{plim} N^{-1} Z'Z = Q_{ZZ} \text{ a finite HxH positive definite matrix}$$

Relevance:

$$\text{plim} N^{-1} Z'X = Q_{ZX} \text{ a finite HxK positive definite matrix}$$

Exogeneity:

$$\text{plim} N^{-1} Z'\varepsilon = 0_{K1}$$

Definition (Instrument properties)

We assume that the H instruments are linearly independent

$$E(Z'Z) \text{ is non singular}$$

Or equivalently rank (E(Z'Z)= H

(1) Exogeneity: They are uncorrelated with the disturbance.

$$E(\varepsilon_j/z_j)= 0_{Nx1} \Longrightarrow E(\varepsilon_j z_j)= 0$$

can expressed as an **orthogonality condition** or **moment condition**

$$E\left(\underset{(H,1)}{z_j}\underset{(1,1)}{\left(y_j - x_j^{'}\beta\right)}\right) = \underset{(H,1)}{0}$$

So, we have H equations and K unknown parameters

Definition (Identification). The system is identified if there exists a unique vector β such that:

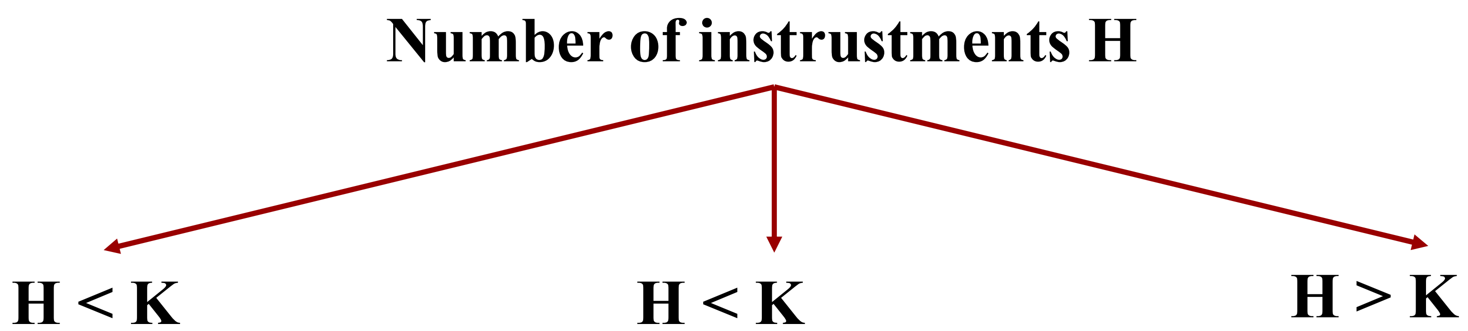
$$E\left(\underset{(H,1)}{z_j}\underset{(1,1)}{\left(y_j - x_j^{'}\beta\right)}\right) = \underset{(H,1)}{0}$$

where $z_j = (z_{j1}..z_{jH})'$. For that, we have the following conditions:

(1) If H < K the model is not identifed.

(2) If H = K the model is just-identifed.

(3) If H > K the model is over-identifed.



5.3.3 Motivation of the IV estimator

By definition of the instruments:

$$p\lim \frac{1}{N} Z'\varepsilon = p\lim \frac{1}{N} Z'\left(y - X\beta\right) = 0_{K\times 1}$$

so we have

$$p\lim \frac{1}{N} Z'y = \left(p\lim \frac{1}{N} Z'X \right)\beta$$

or *equivalently*

$$\beta = \left(p\lim \frac{1}{N} Z'X \right)^{-1} p\lim \frac{1}{N} Z'y$$

If H = K, the Instrumental Variable (IV) estimator $\widehat{\beta}_{IV}$ of parameters β is defined as to be:

$$\widehat{\beta}_{IV} = \left(Z'X\right)^{-1} Z'y$$

Proof

$$\hat{\beta}_{IV} = (Z'X)^{-1} Z'y = (Z'X)^{-1} Z'(X\beta + \varepsilon) = \beta + (Z'X)^{-1}(Z'\varepsilon)$$

$$E\left(\hat{\beta}_{IV}\right) = \beta + \left(\frac{1}{N} Z'X\right)^{-1} \left(\frac{1}{N} Z'\varepsilon\right)$$

so we have

$$p\lim \hat{\beta}_{IV} = \beta + \left(p\lim \frac{1}{N} Z'X\right)^{-1} \left(p\lim \frac{1}{N} Z'\varepsilon\right)$$

Under the assumption of exogeneity of the instruments

$$p\lim \frac{1}{N} Z'\varepsilon = p\lim \frac{1}{N} Z'(y - X\beta) = 0_{K\times 1}$$

so we have

$$p\lim \hat{\beta}_{IV} = \beta$$

5.3.4 Instrumental Variable Estimation

The Instrumental Variable (IV) approach was first proposed by Anderson and Hsiao (1982).

Consider a dynamic panel data model with random individual effects

$$y_{it} = \gamma y_{it-1} + \beta' x_{it} + \alpha_i^* + u_{it}$$

- α^*_i is assumed to be random
- x_{it} is a vector of K_1 time-varying explanatory variables,
- β is a vector of K_1 vector of parameters for the time-varying explanatory variables

5.3.4 Instrumental Variable Estimation (cont.)

$$y_{it} = \gamma y_{it-1} + \beta'_{it} x_{it} + \alpha_i^* + u_{it}$$

Assumption. We assume that the component error term $\varepsilon_{it} = \alpha_i^* + u_{it}$

Remark. If the vector α_i^* includes a constant term, the associated parameter can be interpreted as the **mean** of the (random) individual effects

$$\alpha_i^* = \alpha_0 + \alpha_i \;\; ; E(\alpha_i) = 0$$

About α_i,

$$E(\alpha_i) = 0, \quad ,V(\alpha_i) = E(\alpha_i^2) = \sigma_\alpha^2, \quad ,E(\alpha_i x_{it}) = 0, \quad ,E(\alpha_i \alpha_j) = 0$$

About u_{it},

$$E(u_{it}) = 0, \quad ,V(u_{it}) = E(u_{it}^2) = \sigma_u^2, \quad ,E(u_{it} u_{js}) = 0 \text{ for } i \neq j \text{ and } t \neq s$$

$$E(\alpha_i u_{it}) = 0$$

5.3.4 Instrumental Variable Estimation (cont.)

$$y_{it} = \gamma y_{it-1} + \beta' x_{it} + \alpha_i^* + u_{it}$$

Step 1. first difference transformation

Step 2. choice of the instrutments and IV estimation of γ and β

Step 3. estimation of

Step 4. estimation of the variances σ^2_α and σ^2_u

***

Step 1. first difference transformation

Taking the first difference of the model, we obtain for t = 2, ..,T.

$$y_{it} = \gamma y_{it-1} + \beta' x_{it} + \alpha_i^* + u_{it}$$

$$y_{it-1} = \gamma y_{it-2} + \beta'_{it} x_{it-1} + \alpha_i^* + u_{it-1}$$

$$(y_{it} - y_{it-1}) = \gamma(y_{it-1} - y_{it-2}) + \beta'(x_{it} - x_{it-1}) + u_{it} - u_{it-1} \quad (5.4)$$

- The first difference transformation leads to "lost" one observation
- But, it allows to eliminate the individual effects (as the Within transformation).

Step 2. choice of the instrutments and IV estimation

$$(y_{it} - y_{it-1}) = \gamma(y_{it-1} - y_{it-2}) + \beta'(x_{it} - x_{it-1}) + u_{it} - u_{it-1}$$

Remark. In the difference equation, however, the errors $(u_{it} - u_{it-1})$ are correlated with the regressor $(y_{it-1} - y_{it-2})$.

Therefore, a valid instrument z_{it} should satisfy

$E(z_{it}(u_{it} - u_{it-1})) = 0$, exogeneity property.

$E(z_{it}(y_{it-1} - y_{it-2})) \neq 0$, relevance property.

Eq. (5.4) simply that no have exogeneous the following way:

$$(y_{it} - y_{it-1}) = \gamma(y_{it-1} - y_{it-2}) + \beta'(x_{it} - x_{it-1}) + u_{it} - u_{it-1}$$

Anderson and Hsiao (1981) propose two valid instruments:

- **First instrustment:** $z_{it} = y_{i,t-2}$

$E(y_{it-2}(u_{it} - u_{it-1})) = 0$, exogeneity property.

$E(y_{it-2}(y_{it-1} - y_{it-2})) \neq 0$, relevance property.

- **Second instrustment:** $z_{it} = y_{i,t-2} - y_{i,t-3}$

$E((y_{i,t-2} - y_{i,t-3})(u_{it} - u_{it-1})) = 0$, exogeneity property.

$E((y_{i,t-2} - y_{i,t-3})(y_{it-1} - y_{it-2})) \neq 0$, relevance property.

$$\underbrace{(y_{it} - y_{it-1})}_{\Delta y_{it}} = \gamma \underbrace{(y_{it-1} - y_{it-2})}_{\Delta y_{it-1}} + \underbrace{u_{it} - u_{it-1}}_{\Delta u_{it}}$$

$$\Leftrightarrow \Delta y_{it} = \gamma \Delta y_{it-1} + \Delta u_{it} \quad (5.5)$$

In Eq. (5.5) the errors Δu_{it} are correlated with Δy_{it-1}

Therefore,

$$E(\Delta y_{it-1} \Delta u_{it}) \neq 0$$

Stacking over time, Eq. (5.5) reduces to

$$\Delta y_i = \gamma \Delta y_{i,-1} + \Delta u_i \quad (5.6)$$

Anderson and Hsiao (1981) recommend instrustmenting for Δy_{it-1} with $z_{it} = y_{i,t-2}$ or $y_{i,t-2} - y_{i,t-3}$ which are uncorrelated with the disturbance in (5.5) but correlated with Δy_{it-1}.

The instrumental variable estimation exploits the following moment condition:

$$E(y'_{i,-2}\,\Delta u_{it}) = 0 \quad (5.7)$$

The sample counterpart of (5.7) is

$$E(y'_{i,-2}\Delta u_{it}) = \sum_{i=1}^{N} y'_{i,-2}\left(\Delta y_i - \hat{\gamma}\Delta y_{i,-1}\right) = 0 \quad (5.8)$$

Therefore, using $y_{i,t-2}$, or $y_{i,-2}$ as an instrument for $y_{i,t-1}$, or $y_{i,-1}$, the IV estimator is

$$\sum_{i=1}^{N} y'_{i,-2}\left(\Delta y_i - \hat{\gamma}\Delta y_{i,-1}\right) = 0$$

$$\Leftrightarrow \hat{\gamma}_{1,IV} = \left(\sum_{i=1}^{N} y'_{i,-2}\Delta y_{i,-1}\right)^{-1}\sum_{i=1}^{N} y'_{i,-2}\Delta y_i = \frac{\sum_{i=1}^{N}\sum_{t=2}^{T} y_{i,t-2}\left(y_{it} - y_{it-1}\right)}{\sum_{i=1}^{N}\sum_{t=2}^{T} y_{i,t-2}\left(y_{it-1} - y_{it-2}\right)} \quad (5.9)$$

Now, Eq. (5.8) could be expressed as

$$\hat{\gamma}_{1,IV} = \left(\sum_{i=1}^{N} y_{i,-2}^{'}\Delta y_{i,-1}\right)^{-1}\sum_{i=1}^{N} y_{i,-2}^{'}\Delta y_{i} = \gamma + \left(\sum_{i=1}^{N} y_{i,-2}^{'}\Delta y_{i,-1}\right)^{-1}\sum_{i=1}^{N} y_{i,-2}^{'}\Delta u_{i}$$

Substituting $y_{it-2} = \gamma_0\sum_{j=0}^{t-3}\gamma_1^j + \alpha_i\sum_{j=0}^{t-3}\gamma_1^j + \gamma_1^{t-2}y_{i0} + \sum_{j=0}^{t-3}\gamma_1^j u_{i,t-2-j}$

We have

$$E\left(\hat{\gamma}_{1,IV}\right) = \gamma$$

In general,

$$(y_{it} - y_{it-1}) = \gamma(y_{it-1} - y_{it-2}) + \beta'(x_{it} - x_{it-1}) + u_{it} - u_{it-1}$$

$$\underbrace{(y_{it} - y_{it-1})}_{\Delta y_{it}} = \gamma\underbrace{(y_{it-1} - y_{it-2})}_{\Delta y_{it-1}} + \beta'\underbrace{(x_{it} - x_{it-1})}_{\Delta x_{it}} + \underbrace{u_{it} - u_{it-1}}_{\Delta u_{it}}$$

$$\Leftrightarrow \Delta y_{it} = \gamma\Delta y_{it-1} + \beta'\Delta x_{it} + \Delta u_{it} \quad (5.10) \quad t = 2,3,\ldots,T$$

$$y_{it} = \gamma y_{it-1} + \beta' x_{it} + u_{it} \quad (t=1)$$

Notes.

- The initial first differences model includes $K_1 + 1$ regressors.
- The regressor $y_{it-1} - y_{it-2}$ is endogeneous.
- The regressor $x_{it} - x_{it-1}$ are assumed to be exogeneous.

Anderson and Hsiao (1982) propose two valid instruments:

$$First\ \ instrustment: z_{it} = \left(y_{it-2} \quad (x_{it} - x_{it-1})' \right)'$$

$$\Delta y_{it} = \gamma \Delta y_{it-1} + \beta' \Delta x_{it} + \Delta u_{it} \quad (5.10)$$

$$Y_i = \begin{pmatrix} \Delta y_{i2} \\ . \\ . \\ \Delta y_{iT} \end{pmatrix}; X_i = \begin{pmatrix} \Delta y_{i1} & \Delta x_{i2} \\ . & . \\ . & . \\ \Delta y_{iT-1} & \Delta x_{iT} \end{pmatrix}; Z_i = \begin{pmatrix} y_{i0} & \Delta x_{i1} \\ . & . \\ . & . \\ y_{iT-2} & \Delta x_{iT} \end{pmatrix}$$

$$\text{Second instrustment}: z_{it} = \left(\left(y_{it-2} - y_{it-3} \right) \quad \left(x_{it} - x_{it-1} \right)' \right)'$$

$$Y_i = \begin{pmatrix} \Delta y_{i3} \\ . \\ . \\ \Delta y_{iT} \end{pmatrix}; X_i = \begin{pmatrix} \Delta y_{i2} & \Delta x_{i3} \\ . & . \\ . & . \\ \Delta y_{iT-1} & \Delta x_{iT} \end{pmatrix}; Z_i = \begin{pmatrix} y_{i1} - y_{i0} & \Delta x_{i3} \\ . & . \\ . & . \\ y_{iT-2} - y_{iT-3} & \Delta x_{iT} \end{pmatrix}$$

$$\Delta y_{it} = \gamma \Delta y_{it-1} + \beta' \Delta x_{it} + \Delta u_{it} \quad (5.10)$$

$$\Leftrightarrow Y_i = \delta X_i + \Delta u_i \quad (5.11)$$

$$Y = \begin{pmatrix} Y_1 \\ . \\ . \\ Y_N \end{pmatrix}; X = \begin{pmatrix} X_1 \\ . \\ . \\ X_N \end{pmatrix}; Z = \begin{pmatrix} Z_1 \\ . \\ . \\ Z_N \end{pmatrix}$$

$$Y = \delta X + \Delta u \quad (5.12)$$

$$\Rightarrow \hat{\delta}_{IV} = \left(Z' X \right)^{-1} Z' Y$$

$$First\ \ instrustment: z_{it} = \left(y_{it-2} \quad \left(x_{it} - x_{it-1} \right)' \right)'$$

$$\hat{\delta}_{IV} = \left(Z'X \right)^{-1} Z'Y$$

$$= \left(\sum_{i=1}^{N} \sum_{t=2}^{T} \begin{pmatrix} \Delta y_{it-1} y_{it-2} & y_{it-2} \left(\Delta x_{it} \right)' \\ \Delta x_{it} y_{it-2} & \Delta x_{it} \left(\Delta x_{it} \right)' \end{pmatrix} \right)^{-1} \left(\sum_{i=1}^{N} \sum_{t=3}^{T} \begin{pmatrix} y_{i,t-2} \\ \Delta x_{it} \end{pmatrix} \Delta y_{it} \right)$$

$$\text{Sec}\,ond\ \ instrustment: z_{it} = \left(\left(y_{it-2} - y_{it-3} \right) \quad \left(x_{it} - x_{it-1} \right)' \right)'$$

$$\hat{\delta}_{IV} = \left(Z'X \right)^{-1} Z'Y$$

$$= \left(\sum_{i=1}^{N} \sum_{t=3}^{T} \begin{pmatrix} \Delta y_{it-1} \Delta y_{it-2} & \Delta y_{it-2} \left(\Delta x_{it} \right)' \\ \Delta x_{it} \Delta y_{it-2} & \Delta x_{it} \left(\Delta x_{it} \right)' \end{pmatrix} \right)^{-1} \left(\sum_{i=1}^{N} \sum_{t=3}^{T} \begin{pmatrix} \Delta y_{i,t-2} \\ \Delta x_{it} \end{pmatrix} \Delta y_{it} \right)$$

Remarks.

- The first estimator (with $z_{it} = y_{i,t-2}$) has an advantage over the second one (with $z_{it} = y_{i,t-2} - y_{it-3}$) in that the minimum number of time periods required is two, whereas the first one requires $T \geq 3$.
- In practice, if $T \geq 3$, the choice between both depends on the correlations between $(y_{i,t-1} - y_{it-3})$ and $y_{i,t-2}$ or $(y_{i,t-2} - y_{it-3})$

Pre Example (3.1) With model

ROAA = f(L.ROAA, HHI, L_A, SIZE, ASSET_GRO, GDP, INF) + ε